Terminology databank based on conceptual relationships

Ericsson unveil INTERDOC at Hanover Fair

A new form of terminology database was presented in March at the Hanover Fair, at the information and telecommunications exhibition known as CeBIT '88, by the Stuttgart-based company Ericsson Vertriebs-Partner.

Called INTERDOC (INtegrated TERminology DOCument management system), it provides a tool for the control of a company's whole information processing needs, and assigns key roles to the translator and terminologist.

INTERDOC represents yet another major step forward for the Stuttgart company, which has been working on computer aided tools for the translator and the documentalist for the last ten years.

It began around the end of 1978 when David Computer, the company subsequently acquired (in 1984) by Ericsson, brought out Intext, a word processing package which could be integrated with data processing, one of the first such on the market.

Among the facilities it offered were font generation (making it possible to design your own character) and a dictionary checking function.

This prompted work on a system designed for translators, who would produce their text on screen, and when they wanted to check the translation of a word, they typed Ü (for Übersetzen), the translation appeared at the bottom of the screen, and could be quickly incorporated into the translation. Where the word was not found, the translator could research it, and enter his finding into the system. This was a form of "Wörterbuch in Textverarbeitung".

This envolved into the CAT (Computer Assisted Translation) system which Ericsson has been offering, fully integrated into INTEXT, for the past two to three years. It is designed to operate on the company's own Eritron 200 series equipment.

Of course the care with which new terms are entered is of great importance, and automatic entry of a finding by a translator is given the status "vorläufig" (provisional) until approved by the company terminologist. But the system is entirely user-controlled, and the users of the system decide who is to give the *imprimatur* of approval. Ericsson themselves do not provide dictionaries, though they may of course refer clients to existing databases of terminology, such as EURODICAUTOM and Team.

Terminology which is entered is given source codes and reliability codes, and can where appropriate be given country-specific (e.g. differentiating between UK and US usage in English, or German, Austrian and Swiss usage in German), and even region-specific, codes.

The Ericsson approach is based on the belief that translators will increasingly use direct keyboarding into word processing, rather than the dictation or manual production of translations, and that dictionary look-up is one of the most time-consuming tasks for the translator, and one that should be as far as possible integrated into the word processing system.

Fully automatic systems are considered by Ericsson as being too limited, and that giving the translator electronic support for the way in which he carries out his work is a more rewarding approach.

The INTEXT word processing program is designed with a translator's needs in mind, and as well as normal features such as global search and replace, and integrated electronic mail, teletext and communications capability, it has a multilingual character set and keyboard, a character font generator, and a screen that displays scientific, mathematical and chemical symbols as well as non-Latin alphabets, such as the Cyrillic alphabet.

There are also facilities for automatic line counting (useful for charging or estimating purposes), a facility to set up tables of contents, footnote processing, indexing, conversion of measurements and currencies, translation job administration, determining fluctuations in translation work flow, and access to outside data bases.

When the translator comes to a word which he wants to look up, he presses the translation key and the system presents the possible translations (numbered where there are more than one) plus additional information entered under that term (such as context, definitions, source). By typing the number of the term chosen, the term is automatically entered into the target text in place of the source language term.
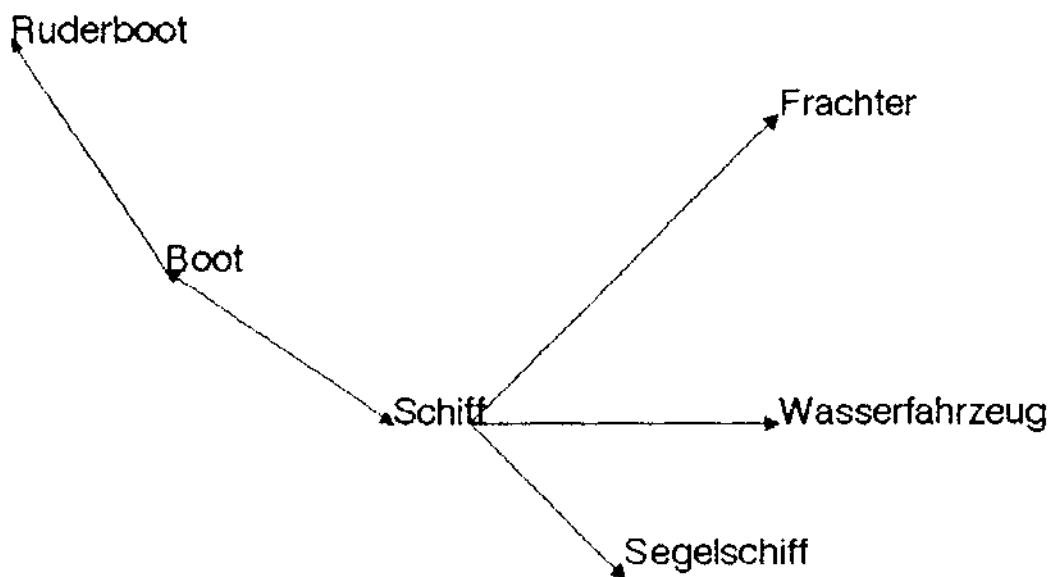
The translator who has researched a not-found term can enter it either there and then, or later in a separate operation. Such new terminology can be entered in any language.

Each term in each language can be classified according to up to six times six different industries or specialisations.

The system is independent of language, since Ericsson provide the facilities, and not the terminology, which is customer-generated. There are plans by the company to add facilities to support more non-Latin scripts, such as those of Japanese, Korean, Thai and Hindi.

INTERDOC

Verbindliches Indexieren

Indexierer   Termbank   Suchende

Schnittstelle zwischen
Indexierer und
Abfragenden

"Schiff" ➡   ⬅ "Boot"

Ruderboot

Frachter

Boot

Schiff ———————➤ Wasserfahrzeug

Segelschiff

Steps are now being taken to make the Ericsson system, which until now has been confined to specific hardware, operate on UNIX, and this should be available towards the end of 1988.

The way the CAT terminology system is constructed led Ericsson to use the term *Knowledge Logistics* for what it was trying to do. It came to consider CAT as "a fully functional first module of a comprehensive solution on the basis of knowledge logistics, since the very concept of the CAT terminology data base ensures integration in a future knowledge logistics system.

Access to terms can be via the lexemes themselves, or by the source reference, by auxiliary information, or by a cross reference. A source reference may be traced back to the actual source, if entered, and thus the term viewed in context.

During the past 3 years working experience with CAT, EVP realised that due to complex market demands, it was necessary to give the CAT-Termbank a totally new structure. "INTERDOC" is the result of those market demands. There is, however, an interface between the "old CAT-Termbank" and INTERDOC's Termbank.

The concept of the "semantic net" is one of the major ways INTERDOC takes this idea further. The terminologist allocates points, on a scale ranging from 0 to 9, to relationships between terms. Thus a "car" will have a fairly close relationship to "transmission" or "bodywork", which in turn would have relationships to "gearbox" or "door". In the illustrated example one could arrive at the term "Boot" directly from "Schiff", and "Ruderboot" directly from "Boot" or indirectly from "Schiff". A "mouse" can be used to "wander" through the semantic net, if desired. In this way it is possible to trace any term which has been entered, and of course the translation, even if one does not know the exact term. This gets over the common problem in terminology searches of different words being used for the same subject (for example, a cursor in German can be a *Lichtmarke*, a *Zeiger*, a *Leuchtmarke*, etc.)

The INTERDOC system is based on the organisation by concepts, and the use of consistent terminology.

The system is sensitive to changes in how terms are understood, which often happens over a period of time. It is also available to different types of users, needing the information for different purposes.

For this reason every department using information should contribute to the building up of the terminology.

The terminology bank is linked to the documentation bank, where full documents or summaries can be stored and accessed by bibliographical data or keywords.

Ericsson Vertriebs Partner have sales offices in Norway, Denmark, Netherlands, France and Switzerland, and CAT with INTERDOC will be introduced in these countries in 1988.

The system is naturally of considerable interest to INFOTERM, the world terminology organisation, based in Vienna, and they have ordered the installation of CAT plus the newly completed (January 1988) preliminary version of INTERDOC for March 1988.

A second preliminary version is expected in April 1988. The definitive availability of INTERDOC is planned for the end of 1989.

On the last working day before he set off for the Hanover Fair Stefan Mahr, export sales manager for Ericsson Vertriebs-Partner, explained to *Language Monthly* the vision his company has of the future for translation and documentation.

The growing importance of terminology is not yet sufficiently realised, but in fact if a large company is to have any control over its documentation it must get its terminology organised. He quotes examples, of three separate departments of a company ordering, individually, sodium hydrosulphide, sodium bisulphide, and sodium sulphydrate until someone realised that the three chemicals were identical. The German Locomotive Standardisation Committee had found that no less than 151 terms were being used for the 14 parts of a stuffing box.

Even before the need for translation is considered, he says, companies are going to have a system for ensuring terminology in their first language is consistent. An efficient information and knowledge management system is becoming imperative. In pharmaceutics, for example, the amount of documentation required for type approval of a new product in America could be as much as a truck load. Any large company which failed to find a way of controlling the mass of information produced would find itself duplicating work already done, perhaps several times over.

Warming to his theme over lunch at a Stuttgart restaurant, he drew two sketches on a table napkin, one showing the usual current situation, where the translator/terminologist is in a corner and only peripherally connected with other departments in a company, the other showing the future situation, whereby the translator/terminologist sits in the middle with all lines of communications, certainly in the documentation field, radiating from him.